Accepted refereed manuscript of: Phillips G, Teixeira H, Poikane S, Salas Herrero F & Kelly MG (2019) Establishing nutrient thresholds in the face of uncertainty and multiple stressors: A comparison of approaches using simulated datasets. *Science of the Total Environment*, 684, pp. 425-433.
DOI: https://doi.org/10.1016/j.scitotenv.2019.05.343

Establishing nutrient thresholds in the face of uncertainty and multiple stressors: a comparison of approaches using simulated data sets

Geoff Philips[1], Heliana Teixeira[2], Sandra Poikane[3], Fuensanta Salas[3] and Martyn G. Kelly[4,5 *]

[1] Biological and Environmental Sciences, University of Stirling, Stirling FK4 9LA, UK

[2] Department of Biology & CESAM, University of Alveira, Campus de Santiago, 3810-193 Aveiro, Portugal.

[3] European Commission Joint Research Centre (JRC), I-21027, Ispra, Italy

[4] Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK.

[5] Department of Geography, Nottingham University, Nottingham NG7 2RD, UK.

* Corresponding author. *E-mail address:* MGKelly@bowburn-consultancy.co.uk (M.G. Kelly)

Abstract

Various methods have been proposed to identify threshold concentrations of nutrients that would support good ecological status, but the performance of these methods and the influence of other stressors on the underlying models have not been fully evaluated. We used synthetic datasets to compare the performance of ordinary least squares, logistic and quantile regression, as well as, categorical methods based on the distribution of nutrient concentrations categorized by biological status. The synthetic datasets used differed in their levels of variation between explanatory and response variables, and were centered at different position along the stressor (nutrient) gradient. In order to evaluate the performance of methods in "multiple stressor" situations, another set of datasets with two stressors were used.Ordinary least squares and logistic regression methods were the most reliable when predicting the threshold concentration when nutrients were the sole stressor; however, both had a tendency to underestimate the threshold when a second stressor was present. In contrast, threshold concentrations produced by categorical methods were strongly influenced by the level of the stressor (nutrient enrichment, in this case) relative to the threshold they were trying to predict (good/moderate in this instance). Although all the methods tested had limitations in the presence of a second stressor, upper quantiles seemed generally appropriate to establish non-precautionary thresholds. For example, upper quantiles may be appropriate when establishing targets for restoration, but not when seeking to minimise deterioration. Selection of an appropriate threshold concentration should also attend to the regulatory regime (i.e. policy requirements and environmental management context) within which it will be used, and the ease of communicating the principles to managers and stakeholders .

Key words

eutrophication, phosphorus, nitrogen, threshold, multiple stressors, Water Framework Directive

1. Introduction

Legislation such as the Water Framework Directive (WFD: European Union, 2000) offers significant opportunities to incorporate ecological knowledge into regulatory mechanisms that ensure sustainable water resources. A key assumption behind such legislation is that, if reasons for deterioration of ecosystems can be identified, then appropriate measures can be put in place to remediate and/or protect against future deterioration. To this end, a large number of metrics summarising the response of the aquatic biota have been developed to meet WFD objectives (Birk et al., 2012) and the position of good and high ecological status boundaries have been harmonised between Member States (Birk et al., 2013; Poikane et al., 2014, 2015). In practice, however, the dynamic nature of ecosystems creates uncertainty in relationships between biology and stressors, with the consequence that predictions of the benefits of remediation lack precision (Moe et al., 2015; Prato et al., 2014). This is widely recognised as a major weakness of WFD implementation (Hering et al., 2010; 2015; Carvalho et al., 2018).

A good understanding of the relationship between biology and a pressure should, in theory, enable a regulator to set threshold concentrations beyond which ecological degradation is likely to occur; however, relationships with stressors such as nutrient enrichment are often weak and confounded by interactions with other stressors (Page et al., 2012; Harris and Heathwaite, 2012; O'Hare et al., 2018; Munn et al., 2018). Consequently, the process of defining thresholds also needs to account for uncertainty in the relationship between biology and pressure.

A number of methods for setting thresholds for nutrient concentrations have been described. Broadly speaking, these fall into two types:

- those that assume a continuous response of both explanatory and response variables, from which a threshold concentration can be inferred using linear regression models. This is the case when ecological status (or other measures of biological condition – e.g. Davies & Jackson, 2006) is defined as a position along a metric scale;

- those that assume a categorical response of one of these variables, allowing threshold concentrations to be inferred using a number of approaches, including binomial logistic regression and methods based on the distribution of pressure values within the class(es) of interest. This is the case when ecological status is expressed as one of a number of classes, but would also be relevant if a particular species or habitat required protection. Free et al. (2016), for example, describe the development of nutrient standards to protect the distinctive *Chara*-dominated communities found in shallow, marl lakes in Ireland, protected under the European Union's Habitats Directive (European Community, 1992).

A range of approaches for calculating threshold concentrations, encompassing both of these strategies, have been described in the literature (Dodds & Oakes, 2004; Free et al., 2016; Hausmann et al., 2016; Poikane et al., 2019) but, as far as we know, there has been no attempt to compare the effectiveness of different methods and, importantly, no systematic consideration of the extent to which the values produced may be confounded in the presence of a second stressor. This is important as the biological response to nutrient enrichment in European freshwaters (Phillips et al. 2018), transitional and coastal waters (Salas et al., 2019) is often distorted by the presence of other stressors, posing difficulties for setting nutrient thresholds that ensure the integrity of aquatic systems. Recent evidences indicate furthermore that nutrient stress occurs in 71% to 98% of multi-stress situations in Europe's surface waters (Nõges et al., 2016). The importance of multiple pressures in shaping communities in aquatic systems is now widely acknowledged (Borja et al., 2011; Nõges et al., 2016; Hering et al., 2015; Feld et al., 2016) although there are, as yet, no definitive approaches to setting protective thresholds for constituents of any pressure or stressor "cocktail". Multiple stressors are likely to interact in different ways and their effects can be difficult to predict, as there is evidence for additive, synergistic and antagonistic effects of multiple stressors in aquatic ecosystems depending on the nature of those stressors and the type of ecosystem (e.g. Jackson et al., 2016; Gieswein et al., 2017; Rodrigues et al., 2018).

Although the WFD requires member states (MS) to establish threshold values for physico-chemical metrics that support good status there is no requirement for these values, unlike the biological metrics, to be harmonised. Given that different methods were used and the inherent uncertainty in relationships, it is not surprising that a wide range of values are now in use across Europe (Phillips and Pitt, 2016). To overcome this and facilitate the use of more uniform threshold values, guidance supported by a statistical toolkit has been produced (Phillips et al. 2018). This encourages the use of a variety of approaches and in this paper we apply these approaches to synthetic datasets, designed to resemble stressor-response relationships between nutrient enrichment and biological community changes, in order to draw out some general lessons on the suitability of different approaches in situations where nutrients are the principal stressor shaping biological communities, and also in the presence of a second stressor. It is not our intention to investigate complex effects. Rather, we use simulated datasets to support identification of data patterns and show sensitivity of commonly used statistical methods to the presence of unmeasured stressors.

2 Materials and methods

2.1 Datasets

In order to make comparisons between the different methods, a series of synthetic data sets were produced. Each data set contained 200 random values of total phosphorus (TP) concentration, a simulated observed Ecological Quality Ratio (EQR) representing overall environmental conditions where only this single stressor influences the observed EQR, and a second simulated EQR where the value was determined by a combination of phosphorus and a second unknown stressor that also had a negative effect on the observed EQR. Apart from the negative effect of both stressors, the synthetic data does not explore the nature of the interaction between the two stressors (i.e. additive, synergistic or antagonistic, *sensu* Pigot et al., 2015) as this is also often unknown in real case scenarios. The only assumption is that where a second stressor is suspected different methods will show different sensitivities to its presence. Likewise, the nutrient thresholds derived from the relationship observed may be more or less reliable or approximate to the "true" value of the measured stressor depending on the method.

Each data set was generated as follows:

1) A normally distributed random set of 200 total phosphorus (TP) concentrations with a known mean and standard deviation was created. The distribution of these data was chosen such that the true EQR would range across the biological gradient from high to moderate status;

2) A "true" EQR was generated from these values, using the regression parameters for a relationship between $\log_{10}$TP and EQR. (parameters taken from the relationship between TP and phytoplankton in lakes used during the Central-Baltic GIG lake intercalibration exercise: Phillips et al., 2014):

$$EQRTrue = -0.62(Log_{10}\,TP) + 1.79. \qquad \text{equation 1}$$

This equation can be re-arranged to determine the "true" TP concentration at the good/moderate boundary EQR, assumed to be 0.6:

$$TP = 10\text{^}((0.6 - 1.79)/-0.62 = 83\ \mu gL^{-1} \qquad \text{equation 2}$$

3) A simulated observed EQR was then created from TP by adding a normally distributed random error term (E), which had a mean of 0 and a known standard deviation (Figure 1a).

$$EQRSimObs = -0.62(Log_{10}\,TP) + 1.79 + E \qquad \text{equation 3}$$

4) Another normally-distributed set of EQR values (EQR2ndPressure) with a fixed mean and standard deviation) was then generated to represent a hypothetical second stressor together with a random probability (0-1) that this second stressor occurs at a particular site.

5) A simulated observed EQR (EQRSimObs 2 pressures) resulting from both TP and the 2nd stressor was calculated by taking the lowest of either the simulated EQR from phosphorus (EQRSimObs) or the 2nd stressor EQR (EQR2ndPressure) where the probability of the second stressor was >0.5. Where the probability of the 2nd stressor was ≤ 0.5, the EQR from phosphorus was used. Scatter plots produced from this approach typically had "wedge-shaped" data clouds, an example of which is shown in Figure 1b.

To assess the effect of different levels of uncertainty and data that span different levels of stress, ten replicate data sets were generated with the same mean TP and error standard deviation. The process was repeated using ten different mean TP values (40, 50, 60, 70, 80 ,90, 100, 110, 120, 130 μgl^{-1}) and 10 different error standard deviations (0.12 – 0.30), representing increased scatter in the true relationship, to finally produce 1000 data sets for each of the single and two-stressor scenarios, an example with mean TP of 50 μl^{-1} and error standard deviation of 0.15 is shown in figure 1.

2.2 Methods for estimating nutrient threshold concentrations

The following methods were used to identify threshold concentrations of TP corresponding to the good/moderate status boundary (assumed to be EQR = 0.6):

Ordinary least-squares regression (OLS): The most obvious approach uses a linear regression between EQR (dependent variable) and log TP (independent variable), with nutrient threshold values determined from the regression parameters.

Logistic regression: An alternative approach that treats ecological status as a categorical variable where a logistic model is fitted between categorical data using a binary response, "biology moderate or worse" = 1 or "biology good or better" = 0 and log TP. Threshold concentrations are determined to be where the probability of being moderate or worse was 0.5. In the case of two stressors an additional value was determined at probability of 0.75

Categorical methods: nutrient concentrations associated with a particular ecological status class (e.g. good ecological status) could also be expressed as a distribution from which an upper quantile might be chosen to indicate a nutrient concentration above which good status was very unlikely to be achieved, or a lower quantile below which good status was very likely to be achieved, so long as nutrients were the main drivers of status. However, the variation inherent in biology-nutrient relationships means that there will be many instances where lower concentrations of nutrients are not associated with good status and vice-versa. The risk of misclassification could, therefore, be reduced by also considering the distribution of nutrient concentrations in the adjacent class (moderate, in this case), where a lower quantile could be adopted to indicate the nutrient concentration below which moderate status was unlikely (and good status was likely to be achieved). Three different approaches were included in these comparisons: average of medians of adjacent classes; average of adjacent quartiles (75^{th} percentile of "good status" and 25^{th} percentile of "moderate status") and the use of the 75^{th} percentile of "good status" alone.

Minimisation of mismatch of classification: An approach that estimates the nutrient threshold value by minimising the mismatch between status (good or better and moderate or worse) for ecological status and the stressor. The method calculates the proportion of records where the biological status is better than the stressor and where it is worse for incremental values of the nutrient threshold. The nutrient threshold value where these two sets of proportions are equal determines the point at which there is the lowest mismatch of classifications. To determine this value a bootstrap approach was used. For each data set 75% of the data were randomly selected and the proportions of mis-classification determined. A loess model was then fitted to these data to determine the nutrient concentration where the mismatch was equal. This was repeated 50 times and the mean nutrient concentration determined.

Linear quantile regression: when the nutrient-biology interactions is confounded by other stressors or environmental variables, the variance around the mean of the response variable is also a function of those explanatory variable(s), leading to wedge-shaped distributions. In such cases, the quantile regression allows different rates of change in the response variable to be predicted along the upper (in the presence of stressors) or lower (in the presence of mitigating factors) boundary of the conditional distribution of the data (Cade and Noon, 2003). The choice of an appropriate quantile to use is somewhat arbitrary, though more extreme values will have a greater potential to be influenced by outliers. We have used the 75^{th} percentile as a compromise that enables upper threshold to be modelled with a reasonable degree of precaution and confidence.

2.3 Comparison of methods

Each of these methods was applied to the synthetic data sets and the predicted good/moderate threshold concentrations (assumed to be EQR = 0.6) for each were recorded for comparison with the "true" threshold concentration defined by equation 2 above. The extent of uncertainty was also recorded using the coefficient of determination (r^2) from the regression between EQRSimObs and TP, these values were categorised into 5 levels (5 ≥0.6, 4 ≥0.5, 3 ≥0.4, 2 ≥ 0.3, 1 <0.3) to allow the effect of scatter to be determined. To assess the influence of the data distribution along the stressor gradient, results for each of the data sets categorised by mean TP were compared. When applying the methods to real datasets, the threshold nutrient at the good to moderate EQR boundary would be unknown so it would be impossible to determine how a derived threshold relates to the average TP. In such cases, the mean EQR could be used so we present results for the synthetic datasets using the true EQR value determined from the mean TP of the data set using equation 1.

All analyses were performed using R statistical software (R Development Core Team, 2016). Base statistics were used for all methods except linear quantile regression, which was fitted using the rq function from quantreg package (Koenker and Hallock 2001).   Original computer code is available from the first author on request.

3 Results

3.1 Average differences

A comparison of the range of predicted TP threshold values for the Good-Moderate boundary shows that ordinary least squares (OLS) regression and binary logistic regression at a probability of 0.5 predicted the smallest range of values (c ± 5 µg TP L^{-1}, Figures 2 and 3). The variability of the categorical methods was substantially higher (typically ± 15 µg TP L^{-1}), while the minimisation of mismatch method predicted a range of values that lie between the regression and categorical methods (±8 µg TP L^{-1}). When TP was treated as a single stressor all methods, except the 75th percentile of the TP concentration in sites with good- biological status, predicted values that were centred around the true threshold value (83 µg l^{-1} – see equation 2). The 75th percentile predicted significantly higher values than the other methods ($F = 163$ $p < 0.001$, Figure 2).

When a second stressor was present (Figure 3) the predicted range of values did not change, but both linear and logistic regression (at a probability of 0.5), underestimated the true threshold value by 36 µg TP L^{-1}, suggesting that these methods are not appropriate under such circumstances. In contrast the categorical methods were less influenced; the two averaging approaches (median and adjacent quartiles) slightly underestimating (-5 µg TP L^{-1}), with the upper 75 percentile closer to the true mean (+12 µg TP L^{-1}). The minimisation of mismatch method also underestimated the true threshold, although less so than was the case for the regression methods (-20 µg TP L^{-1}). Quantile regression, using the 0.75 quantile and the logistic regression using a probability of 0.75 provided nutrient threshold estimates for the good-moderate boundary that were higher than the true threshold (+26 µg TP L^{-1} and +34 µg TP L^{-1}).

3.2 Influence of position of data cloud along stressor gradient and at different levels of variability

For a single stressor neither OLS nor logistic regression (using a probability of 0.5) methods were significantly influenced by either their uncertainty or position on the stressor gradient (Figure 4 & Table 1). In contrast, all of the categorical methods and the mismatch approach were significantly influenced by the level of stressor, under-estimating the true threshold at low exposures (i.e. predicting a lower nutrient threshold thus more stringent than necessary) and over-estimating at high pressures (i.e. predicting a higher nutrient threshold thus more relaxed than that required). The average of adjacent quartiles and the 75th percentile of TP in good-moderate status were the most sensitive to uncertainty, the minimisation of mismatch the least, but all had a significant interaction term showing an increasing effect of uncertainty as the stressor level increased. Where a second stressor was present, a similar pattern was seen (Figure 5 and Table 1), although the effects of uncertainty and stressor levels were slightly higher. For example, both OLS and logistic regressions ($p = 0.5$) predicted boundary values that were significantly affected by both variability and their position on the stressor gradient when two stressors influenced ecological status, although the effect was much smaller than for the other methods. Logistic regression predictions using $p = 0.75$ were particularly influenced by position on the stressor gradient, overpredicting the true threshold at low stressor levels. The predictions using quantile regression ($p = 0.75$) were not significantly influenced by stressor level, but were by variability, with higher predictions at high levels of uncertainty.

4 Discussion

Where a single stressor dominates the response of biology, linear regression or binary logistic regression are the most reliable approaches. Neither are substantially influenced by the mean of the data set and both are only slightly influenced by scatter in the data. This is unsurprising given that the data were generated with normally distributed errors and thus conform to the requirements of regression models.

Any attempt to develop nutrient thresholds in freshwaters or coastal waters, however, also needs to be aware that nutrients rarely act in isolation (Vinebrooke et al., 2004; Wagenhoff et al., 2011; Piggott et al., 2015; Gunderson et al., 2016), particularly in rivers and estuaries, and our analyses indicate how interactions with a second stressor can confound the face-value relationship between biology and the stressor of interest. Consideration of the complex relationships between the ecological response and stressors acting simultaneously is essential to decide management actions, because of non-linear and interactive effects of stressors (Brown et al.,2013)

In these situations, neither linear nor logistic regressions are appropriate as the confounding effect of the second stressor will result in the under estimation of nutrient thresholds. Such data show heteroscedasticity, with decreasing variance as stressor levels increase, caused by the 2^{nd} stressor overriding the otherwise low influence of nutrient effects. The categorical approaches initially appear to be less influenced by this problem, as on average they make predictions that are clustered around the true mean. However, unlike the regression methods, they are much more sensitive to the position of the data cloud on the stressor gradient. If the data are clustered around the boundary being predicted (the good/moderate boundary, EQR = 0.6, in our study), they are the least sensitive to the effect of a 2^{nd} pressure. However if the data are centred below or above the boundary of interest, they are likely to under- and overpredict, respectively, with the threshold error increasing as uncertainty increases. The least influenced was the minimisation of mismatch method, but all the approaches, other than those seeking to describe the behaviour of the upper distributions of the data, are likely to underestimate threshold values due to the influence of other stressors.

The best solution to this problem would be to develop a more complex model that could account for additional pressures; however, a lack of reliable data and the complexity of modelling make this impractical (Feld et al., 2016 Duarte et al., 2009). Whilst the combined effect of multiple stressors was previously assumed to be additive , this is not always the case in ecological systems, where antagonistic and synergistic interactions may dominate (e.g. Crain et al., 2008; Jackson et al., 2016; Gieswein et al., 2017; Munn et al., 2018; Rodrigues et al., 2018).

An alternative approach would be to fit an upper quantile, which identifies an upper surface to the relationship between EQR and nutrient concentration. The problem with this approach is that it needs to consider the uncertainty in the relationship between EQR and phosphorus. Our simulations show that quantile regression predicts higher values as uncertainty increases. As the uncertainty of the true relationship between nutrient and EQR decreases, a clearer upper boundary emerges, with the upper quantile that is modelled to determine a threshold value approaching, or better reflecting, the true effect of the single stressor. On the other hand, it does indicate the highest values of a physicochemical parameter that is consistent with good status (Müller et al., 2017). Beyond this point, nutrients are likely to exert an effect regardless of the presence of other stressors.

There is no ‘correct’ quantile, and one should inspect the distribution of quantiles within the particular range of interest (Koenker and Hallock, 2001). Higher quantiles offer greater chance that the true response to nutrient stressor is being captured but with the risk that the regression line is anchored by fewer, and more extreme, records at any level of pressure (Koenker, 2011). This problem is particularly acute with small datasets. In practice, the 75th percentile offers a balance between precaution and statistical robustness when dealing with medium-size datasets, although our simulations suggest that even this value may be too high, over predicting at all levels of pressure.

Choosing an upper probability value with logistic regression is a similar approach, potentially allowing threshold values to be determined when a second pressure is present despite any confounding effects. However, again it is difficult to determine the appropriate probability to use. The selection of probability should obey fit-for-purpose criteria, for which several classification measures exist that can be used as support (Fielding and Bell, 1997).

Taylor et al. (2018) advocate a combination of spatial, temporal and experimental approaches in order to characterise the response of biota to nutrient enrichment whilst, at the same time, recognising that comprehensive study designs can become prohibitively expensive. Their study was limited to a single group of biota, diatoms, whilst we would advocate the examination of the response of different ecosystem components to enhance the insights (Robertson et al, 2006). Teichert et al. (2016), by contrast, used a random forest algorithms to detect the dominant stressors in estuaries. At the heart of these approaches, however, lie datasets that capture the spatial and/or temporal variation in assemblages along a strong nutrient gradient and it is also important that statistical analysis of such datasets are both robust and easy to communicate to non-specialist managers and stakeholders.

In practice, however, such approaches are a necessary element when developing such thresholds because they offer the most straightforward means of capturing the range of uncertainty associated with the water body type under investigation. It is, however, important to validate thresholds using independent sources of evidence. The use of experimental systems (Bowes et al., 2012; McCall et al., 2017; Taylor et al., 2018;) is one means of doing this, but other options are also available (e.g. Free et al., 2016).

5 Conclusions

Our simulations suggest that, where there is a strong stressor-response relationship between nutrients and ecological status, any of the tested modelling methods, with the exception of the threshold derived from the 75th percentile of nutrient concentration in sites with good ecological status, are likely to give reliable estimates of nutrient concentrations that are associated with the ecological good-moderate boundary. Of these, OLS or logistic regression are the most reliable, while the minimisation of mismatch method is perhaps the easiest to communicate to managers. This is likely to be the situation for lakes where the dominance of the algal response to nutrients is clear.

In rivers, estuaries and coastal waters however, multiple stressors are common; here the assumed robust regression approaches may be strongly influenced by stressors other than nutrients and there is a risk that threshold values that are lower than needed may be generated, in effect penalizing nutrients for impacts caused by other stressors. Such situations can be identified from wedge shaped scatter plots and from plots of model residuals and it is important that these are carefully considered before the results of modelling are translated to regulatory regimes.

Where there is evidence of multiple stressors, quantile regression or the use of logistic regression with nutrient threshold concentrations are determined using a quantile or a probability greater than 0.5 have potential. However, the selection of an appropriate quantile remains an unresolved issue. Supporting chemical element thresholds values determined for different EQR categories are unlikely to be precautionary as, by their nature, they seek to minimise false positives, i.e. effect detection when there is no effect. Such boundaries may be appropriate when establishing targets for restoration, but less so when seeking to minimise deterioration.

Eutrophication is a complex issue (Dodds, 2006; O'Hare et al., 2018) but, for strategic planning and high-level overviews, there are still benefits in knowing threshold values beyond which consequences can be expected. Understanding the challenges involved in deriving such targets does, at least, enable regulators to interpret results, and combine various strands of evidence in to make robust decisions.

Declaration of interests

None

Acknowledgements

This work was supported by the European Commission (DG-ENV) Joint Research Centre (JRC). We are grateful to two anonymous reviewers for helpful comments on the manuscript

References

Birk, S., Bonne, W., Borja, Á., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18, 31–41. doi:10.1016/j.ecolind.2011.10.009

Birk, S., Willby, N.J., Kelly, M.G., Bonne, W., Borja, A., Poikane, S., van de Bund, W., 2013. Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. Science of The Total Environment 454–455, 490–499. doi:10.1016/j.scitotenv.2013.03.037

Borja, Á., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Elliott, M., Garmendia, J.M., Marques, J.C., Mazik, K., Muxika, I., Neto, J.M., Norling, K., Rodríguez, J.G., Rosati, I., Rygg, B., Teixeira, H., Trayanova, A., 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Marine Pollution Bulletin 62, 499–513. doi:10.1016/j.marpolbul.2010.12.009

Bowes, M.J., Ings, N.L., McCall, S.J., Warwick, A., Barrett, C., Wickham, H.D., Harman, S.A., Armstrong, L.K., Scarlett, P.M., Roberts, C., Lehmann, K., Singer, A.C., 2012. Nutrient and light limitation of periphyton in the River Thames: Implications for catchment management. Science of the Total Environment 434, 201-212. doi:10.1016/j.scitotenv.2011.09.082

Brown, C.J., Saunders, M.I., Possingham, H.P., Richardson, A.J., 2013. Managing for interactions between local and global stressors of ecosystems. PLoS One 8, e65765.

Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1, 412-420. doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2

Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk, S., Blackstock, K.L., Borics, G., Borja, A., Feld, C.K., Ferreira, M.T., Globevnik, L., Grizzetti, B., Hendry, S., Hering,D., Kelly, M., Langaas, S., Meissnern, K., Panagopoulos, Y., Penning, E., Rouillard, J., Sabater, S., Schmedtje, U., Spears, B.M., van de Bund, W., Solheim, A.L., 2018. Protecting and restoring Europe's waters: an analysis of the future development needs of the Water Framework Directive. Science of the Total Environment (in press) doi.org/10.1016/j.scitotenv.2018.12.255

Crain, C.M., Kroeker, K., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecol. Lett. 11, 1304–1315

Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16, 1251-1266. doi:10.1890/1051-0761(2006)016[1251:TBCGAD]2.0.CO;2

Dodds, W.K., 2006. Eutrophication and trophic state in rivers and streams. Limnology and Oceanography 51, 671-680. doi:10.4319/lo.2006.51.1_part_2.0671

Dodds, W.K., Oakes, R.M., 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography Methods 2, 333-341. doi:10.4319/lom.2004.2.333

Duarte, C. M., Conley, D. J., Carstensen, J., & Sánchez-Camacho, M., 2009. Return to Neverland: shifting baselines affect eutrophication restoration targets. Estuaries and Coasts 32(1), 29-36. doi:10.1007/s12237-008-9111-2

European Community (1992). Council Directive 92 /43 /EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities L 206: 7-50.

European Union (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23rd October 2000 Establishing a Framework for Community Action in the Field of Water Policy Official Journal of the European Communities, European Commission, Brussels (2000) (22 December, L 327/1)

Feld, C.K., Segurado, P., Gutiérrez-Cánovas, C., 2016. Analysing the impact of multiple stressors in aquatic biomonitoring data: a cookbook with applications in R. Science of the Total Environment 573, 1320-1339. doi:10.1016/j.scitotenv.2016.06.243

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, 38–49. doi:10.1017/S0376892997000088

Free, G., Tierney, D., Little, R., Kelly, F., Kennedy, B., Plant, C., Trodd, W., Wynne, C., Caroni, R., Byrne, C., 2016. Lake ecological assessment metrics in Ireland: relationships with phosphorus and typology parameters and the implications for setting nutrient standards. Biology and Environment: Proceedings of the Royal Irish Academy 116B, 191-204. doi:10.3318/bioe.2016.18

Gieswein A., Hering D. & Feld C.K. (2017). Additive effects prevail: The response of biota to multiple stressors in an intensively monitored watershed. Science of The Total Environment 593-594, 27-35. http://dx.doi.org/10.1016/j.scitotenv.2017.03.116

Gunderson, A. R., Armstrong, E. J., Stillman, J. H.,2016. Multiple stressors in a changing world: the need for an improved perspective on physiological responses to the dynamic marine environment. Annual Review of Marine Science, 357-378. doi:10.1146/annurev-marine-122414-033953

Harris, G.P., Heathwaite, A.L., 2012. Why is achieving good ecological outcomes in rivers so difficult? Freshwater Biology 57, 91-107. doi:10.1111/j.1365-2427.2011.02640.x

Hausmann, S., Charles, D.F., Gerritsen, J., Belton, T.J., 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Science of the Total Environment 562, 914-927. doi:10.1016/j.scitotenv.2016.03.173

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K., Moe, J., Pont, D., Lyche-Solheim, A., v. de Bund, W., 2010. The european water framework directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment 408 (19), 4007-4019. doi:10.1016/j.scitotenv.2010.05.031

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A.C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A.L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M., Birk, S., 2015. Managing aquatic ecosystems and water resources under multiple stress — An introduction to the MARS project. Science of the Total Environment 503-504, 10-21. doi:10.1016/j.scitotenv.2014.06.106

Jackson, M. C., Loewen, C. J. G., Vinebrooke, R. D., Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. Global Change Biology 22, 180–189, http://dx.doi.org/10.1111/gcb.13028

Koenker, R., Hallock, K.F., 2001. Quantile regression. Journal of Economic Perspectives 15(4), 143-156. doi:10.1257/jep.15.4.143

Koenker, R., 2011. Additive models for quantile regression: Model selection and confidence bandaids. Brazilian Journal of Probability and Statistics 25(3), 239-262. doi:10.1214/10-BJPS131

McCall, S.J., Hale., M.S, Smith, J.T., Read, D.S., Bowes, M.J., 2017. Impacts of phosphorus concentration and light intensity on river periphyton biomass and community structure. Hydrobiologia 792: 315-330. doi: 10.1007/s10750-016-3067-1

Moe, S.J., Lyche-Solheim, A., Soszka, H., Gołub, M., Hutorowicz, A., Kolada, A., Picinska-Fałtynowicz, J., Białokoz, W., 2015. Integrated assessment of ecological status and misclassification of lakes: The role of uncertainty and index combination rules. Ecological Indicators 48, 605-615. doi:10.1016/j.ecolind.2014.08.018

Müller, A., Halle, M., Bellack, E., 2017. Schwellenwerte und biologische Indikatoren für physikalischchemische Parameter in Fließgewässern. (Threshold values and biological indicators for physicochemical parameters in running waters). Wasser und Abfell 06/2017, 25-30. doi:10.1007/s35152-017-0068-y

Munn, M. D., Waite, I., Konrad, C. P. (2018). Assessing the influence of multiple stressors on stream diatom metrics in the upper Midwest, USA. Ecological Indicators 85, 1239-1248. https://doi.org/10.1016/j.ecolind.2017.09.005

Nõges, P., Argillier, C., Borja, Á., Garmendia, J.M., Hanganu, J., Kodeš, V., Pletterbauer, F., Sagouis, A., Birk, S., 2016. Quantified biotic and abiotic responses to multiple stress in freshwater, marine and ground waters. Science of the Total Environment 540, 43-52. doi:10.1016/j.scitotenv.2015.06.045

O'Hare, M.T., Baattrup-Pedersen, A., Baumgarte, I., Freeman, A., Gunn, I.D.M., Lázár, A.N., Wade, A.J., Bowes, M.J., 2018. Responses of aquatic plants to eutrophication in rivers: a revised conceptual model. Frontiers in Plant Science 9, 451. doi:10.3389/fpls.2018.00451

Page, T., Heathwaite, A.L., Moss, B., Reynolds, C., Beven, K.J., Pope, L., Willows, R., 2012. Managing the impacts of nutrient enrichment on river systems: dealing with complex uncertainties in risk analysis. Freshwater Biology 57, 108-123. doi:10.1111/j.1365-2427.2012.02756.x

Phillips, G., Pitt, J-A., 2016. A comparison of European freshwater nutrient boundaries used for the Water Framework Directive: A report to WG ECOSTAT. University College London. Available via: https://circabc.europa.eu/w/browse/58a2363a-c5f1-442f-89aa-5cec96ba52d7

Phillips G, Free G, Karottki I, Laplace-Treyture C, Maileht K, Mischke, U., Ott, I., Pasztaleniec, A., Portielje, R, Søndergaard, M., Trodd, W. & Van Wichelen, J., 2014. Water Framework Directive Intercalibration Technical Report: Central Baltic Lake Phytoplankton ecological assessment methods. Report EUR 26508 EN. Luxembourg: Publications Office of the European Union. DOI:10.2788/73991

Phillips, G., Kelly, M., Teixeira, H., Salas, F., Free, G., Leujak ,W., Lyche Solheim, A., Várbíró, G., Poikane, S., 2018. Best practice for establishing nutrient concentrations to support good ecological status, EUR 29329 EN, Publications Office of the European Union, Luxembourg. doi:10.2760/84425, JRC112667

Piggott, J.J., Townsend, C.R., Matthaei, C.D., 2015. Reconceptualizing synergism and antagonism among multiple stressors. Ecology and Evolution 5, 1538-1547. doi:10.1002/ece3.1465

Poikane, S., Zampoukas, N., Borja, A., Davies, S.P., van de Bund, W., Birk, S., 2014. Intercalibration of aquatic ecological assessment methods in the European Union: Lessons learned and way forward. Environmental Science & Policy 44, 237-246. doi:10.1016/j.envsci.2014.08.006

Poikane, S., Birk, S., Böhmer, J., Carvalho, L., de Hoyos, C., Gassner, H., Hellsten, S., Kelly, M., Solheim, A.L., Olin, M., Pall, K., Phillips, G., Portielje, R., Ritterbusch, D., Sandin, L., Schartau, A.-K., Solimini, A.G., van den Berg, M., Wolfram, G., van de Bund, W., 2015. A hitchhiker's guide to European lake ecological assessment and intercalibration. Ecological Indicators 52, 533-544. doi:10.1016/j.ecolind.2015.01.005

Poikane, S., Phillips, G., Birk, S., Free, G., Kelly, M., Willby, N., 2019. Deriving nutrient criteria to support "good" ecological status in European lakes: an empirically based approach to linking ecology and management. Science of the Total Environment 650, 2074-2084. doi:10.1016/j.scitotenv.2018.09.350

Prato, S., La Valle, P., De Luca, E., Lattanzi, L., Migliore, G., Morgana, J.G., Munari, C., Nicoletti, L., Izzo, G., Mistri, M., 2014. The ‘‘one-out, all-out’’ principle entails the risk of imposing unnecessary restoration costs: A study case in two Mediterranean coastal lakes. Marine Pollution Bulletin 80, 30-40. doi:10.1016/j.marpolbul.2014.01.054

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

Robertson, D. M., Graczyk, D. J., Garrison, P. J., Wang, L., LaLiberte, G., Bannerman, R., 2006. Nutrient concentrations and their relations to the biotic integrity of wadeable streams in Wisconsin. Professional Paper No. 1722, United States Geological Survey, Reston, Virginia.

Rodrigues, A. C., Machado, A. L., Bordalo, M. D., Saro, L., Simão, F. C., Rocha, R. J., Golovko O., Žlábek, V., Barata, C., Soares, A.M.V.M., Pestana, J. L. (2018). Invasive Species Mediate Insecticide Effects on Community and Ecosystem Functioning. Environmental Science & Technology 52(8), 4889-4900. http://dx.doi.org/10.1021/acs.est.8b00193

Salas, F. Teixeira, H., Poikane, S. (2019). A novel approach for deriving nutrient criteria to support good ecological status: Application to coastal and transitional waters and indications for use. Frontiers in Marine Science. DOI: 10.3389/fmars.2019.00255

Taylor, J.M., Back, J.A., Brooks, B.W., King, R.S., 2018. Spatial, temporal, and experimental: three study design cornerstones for establishing defensible numeric criteria for freshwater ecosystems. Journal of Applied Ecology 55, 2114-2123. doi:10.1111/1365-2664.13150

Teichert. N., Borja, A., Chust, G., Uriarte, A., Lepage, M., 2016. Restoring fish ecological quality in estuaries: Implication of interactive and cumulative effects among anthropogenic stressors. Science of Total Environment 542: 383-393

Vinebrooke, R.D., Cottingham, K.L., Norberg, J., Scheffer, M., Dodson, S.I., Maberly, S.C., Sommer, U., 2004. Impacts of multiple stressors on biodiversity and ecosystem functioning: the role of species co-tolerance. Oikos 104, 451-457. doi:10.1111/j.0030-1299.2004.13255.x

Wagenhoff, A., Townsend, C.R., Phillips, N., Matthaei, C.D., 2011. Subsidy-stress and multiplestressor effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams and rivers. Freshwater Biology 56, 1916-1936. doi:10.1111/j.1365-2427.2011.02619.x

List of Tables

Table 1 Analysis of variance table showing the influence of variability (r^2 category) and position of data cloud (mean TP) on TP thresholds predicted for good ecological status using the different methods applied to synthetic data set. Significant F values shown in bold.

Method	Data set	Variability		Position data cloud (mean TP)		Interaction	
		F	p	F	p	F	p
OLS regression	single stressor	3.5	0.06	0.362	0.548	0.332	0.564
	two stressors	**50.8**	<0.001	**29.4**	<0.001	2.9	0.09
Ave. median	single stressor	3.8	0.051	**7731**	<0.001	**312.2**	<0.001
	two stressors	4.1	0.044	**8572.1**	<0.001	**173.7**	<0.001
Ave. quartile	single stressor	**53.5**	<0.001	**53.5**	<0.001	**6091.7**	<0.001
	two stressors	**117.6**	<0.001	**6807.7**	<0.001	**146.2**	<0.001
75th percentile	single stressor	**578.2**	<0.001	**3149**	<0.001	**239.7**	<0.001
	two stressors	**512**	<0.001	**3062.1**	<0.001	**203.5**	<0.001
Mismatch	single stressor	0.2	0.651	**2178.1**	<0.001	**219.1**	<0.001
	two stressors	0.1	0.803	**4069.3**	<0.001	**48.5**	<0.001
Logistic regression (p=0.5)	single stressor	1.2	0.277	0.059	0.809	0.617	0.432
	two stressors	**279.1**	<0.001	**75.2**	<0.001	0.565	0.452
Logistic regression (p=0.75)	single stressor	**36.5**	<0.001	**113.5**	<0.001	1.6	0.209
Quantile regression (p=0.75)	two stressors	**272.9**	<0.001	1.3	0.251	0.342	0.559

Figures

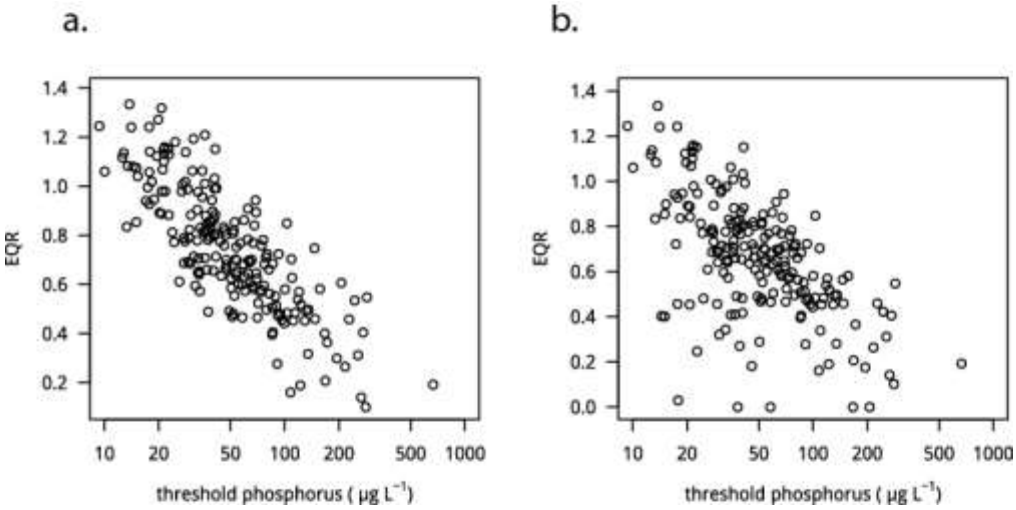

Figure 1. Typical scatter plots showing relationships between simulated EQR and total phosphorus for a) single stressor gradient of phosphorus, b) combined stressor gradient of phosphorus and a second pressure. (Data were for a mean TP of 50 µgl^{-1} and an error standard deviation of 0.15)

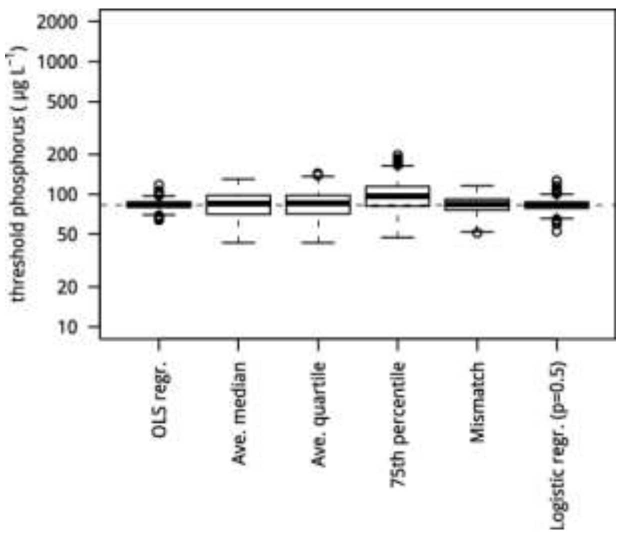

Figure 2. Range of TP concentrations at the good/moderate threshold predicted by the different methods. (In this and subsequent figures the dotted line shows the true threshold concentration (83ug L^{-1})).

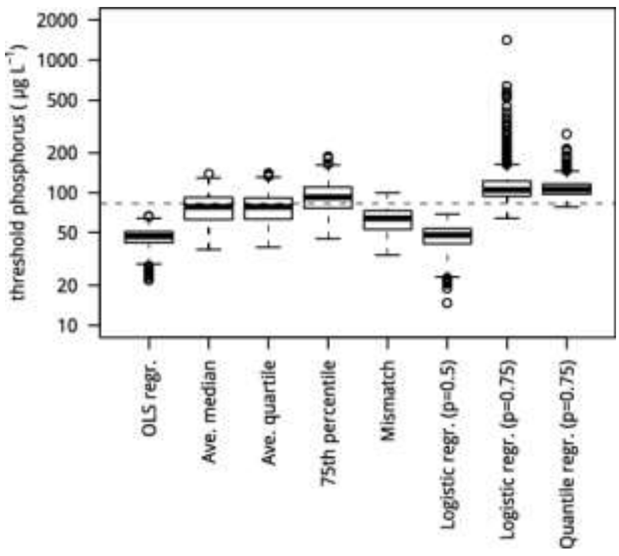

Figure 3. Range of TP concentrations at the good/moderate threshold predicted in the presence of a second pressure ("wedge-shaped" data).

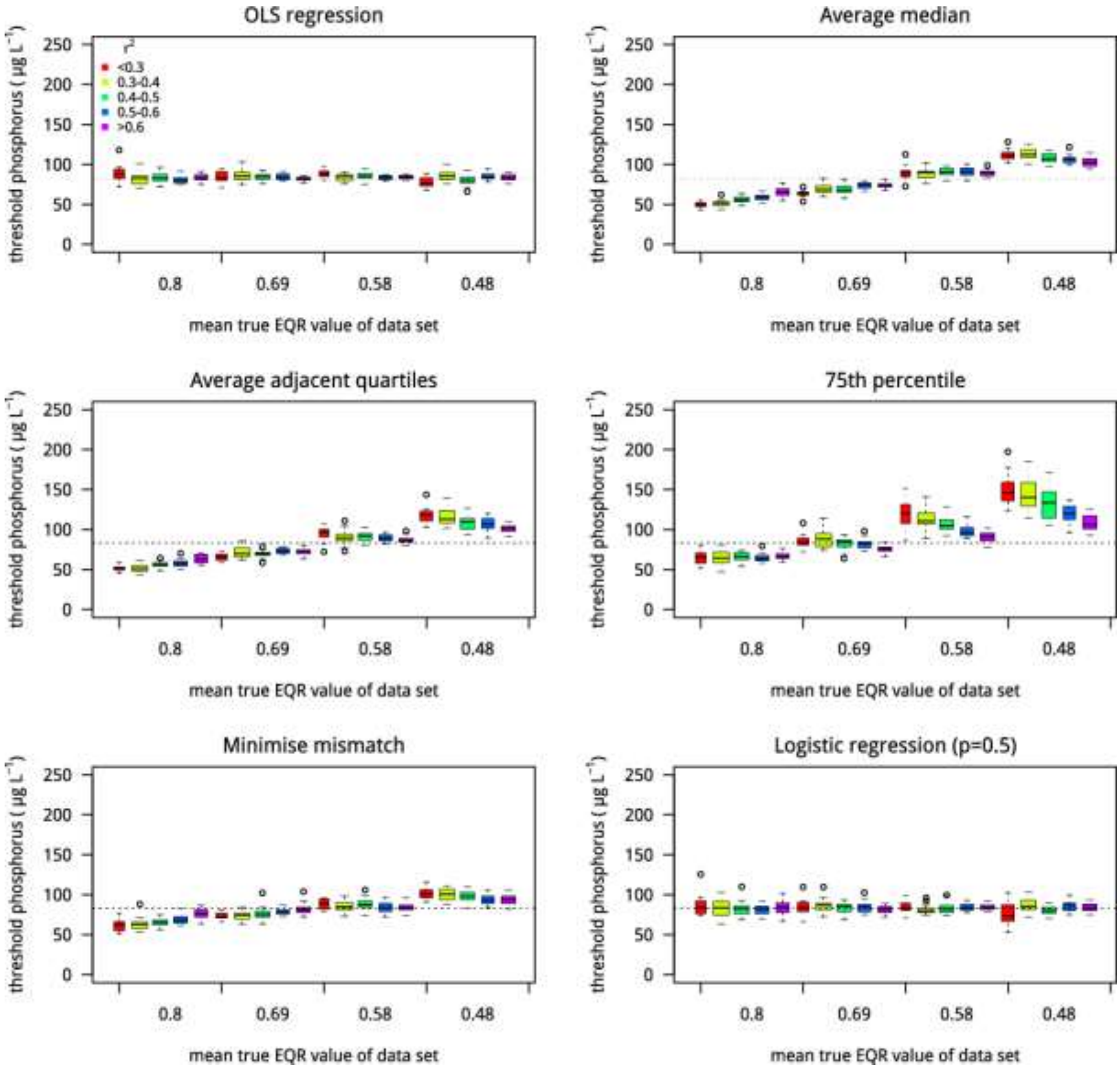

Figure 4 Range of TP concentrations at the good/moderate threshold predicted by each of the methods using simulated data with a single stressor (TP). Boxes grouped by position of the data cloud, characterised by the true EQR calculated using equation 1 from the mean TP of data set and arranged (coloured) by variability of the relationship between the simulated TP and EQR (r^2). (for clarity only 4 of the 10 different stressor levels are shown (40,60,90,130 μgL^{-1})). Dashed line represents the true mean phosphorus threshold.

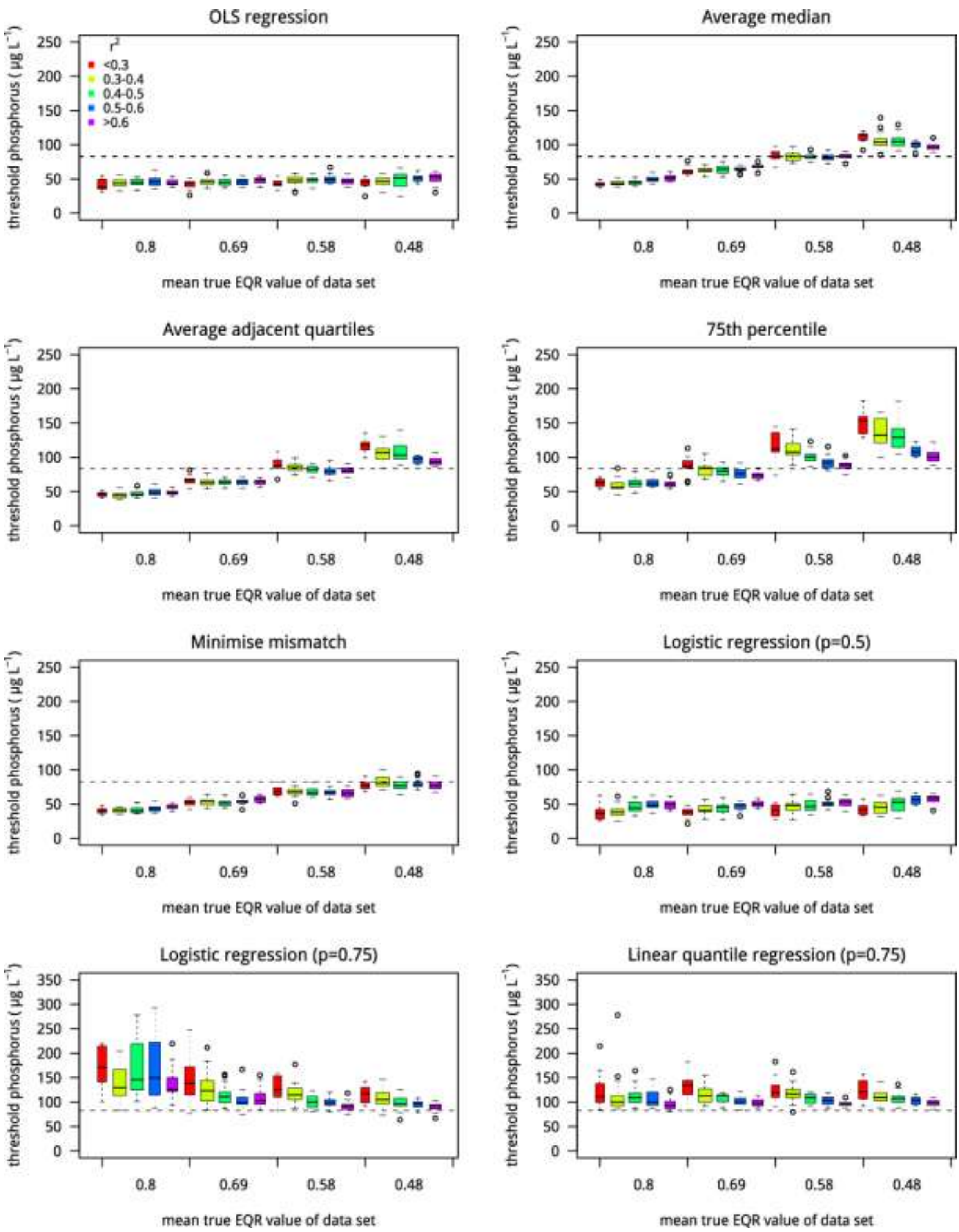

Figure 5 Range of TP concentrations at the good/moderate threshold predicted by each of the methods using simulated data with a stressor (TP) and an additional second unknown stressor. Boxes grouped by position of the data cloud, characterised by the true EQR calculated using

equation 1 from the mean TP of data set and arranged (coloured) by variability of the relationship between the simulated TP and EQR (r^2). (for clarity only 4 of the 10 different pressure categories are shown (40,60,90,130 μgL^{-1})). Dashed line represents the true mean phosphorus threshold.